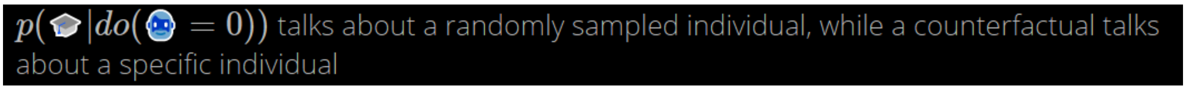
Counterfactuals (Judea Peral's definition)

- *Given that Alice did not get promoted in her job, and given that she is a woman, and given everything else we can observe about her circumstances and performance, what is the probability of her getting a promotion if she was a man instead?*
- The point of asking this question: what's the degree of being a women is directly responsible for the observed outcome (no promotion)?
- Another weird thing to note about this counterfactual is that the intervention (Alice's gender magically changing to male) is not something we could ever implement or experiment with in practice. (In econometrics, this problem would be not well-posed since you cannot set the treatment in this case).
- *Given that I have a beard, and that I have a PhD degree, and everything else we know about me, with what probability would I have obtained a PhD degree, had I never grown a beard.*
- Since we do not expect beard to be causally related to getting Phd, the probability that I would have obtained a PhD had I not grown a beard should be close to 1.
- So counterfactual in Peral's definition is NOT a hypothetical group of people who did not received treatment, but are otherwise exactly the same as the people who received treatment; rather it is the probability of an outcome (Y=1) given that there is no treatment, with everything else staying the same.

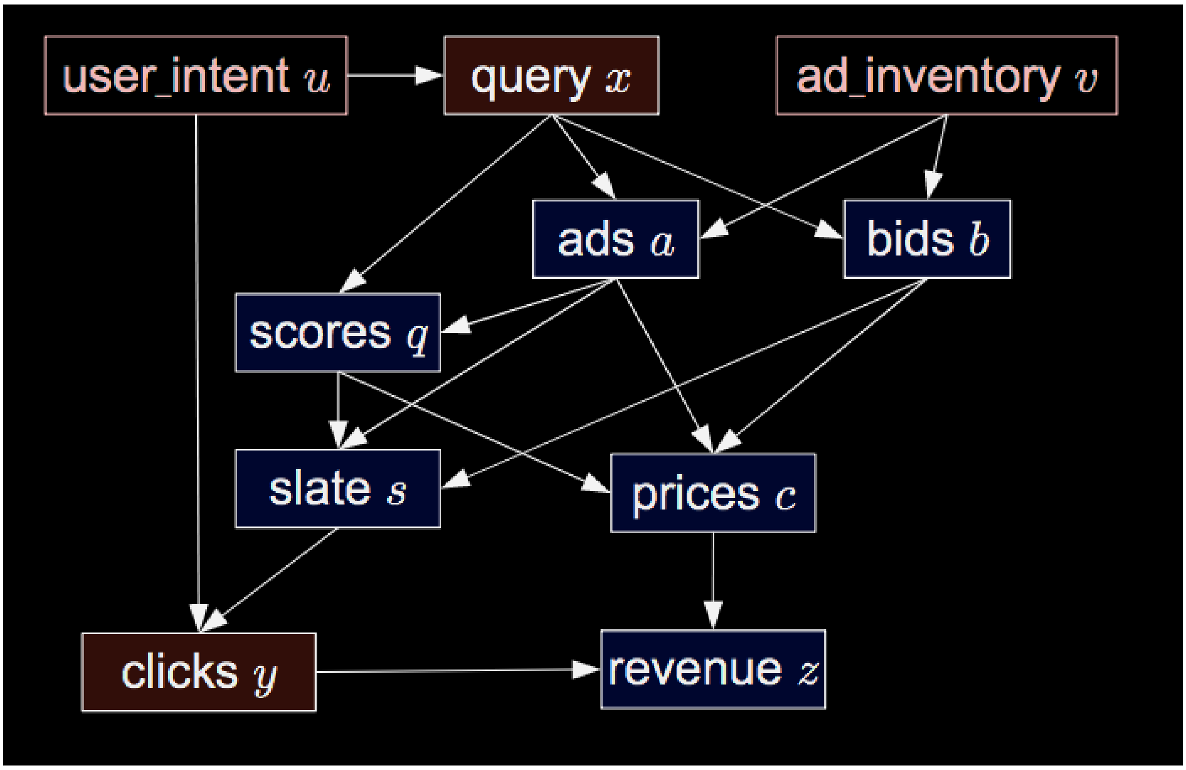


$p($🎓$|do($🧔$=0))$ talks about a randomly sampled individual, while a counterfactual talks about a specific individual

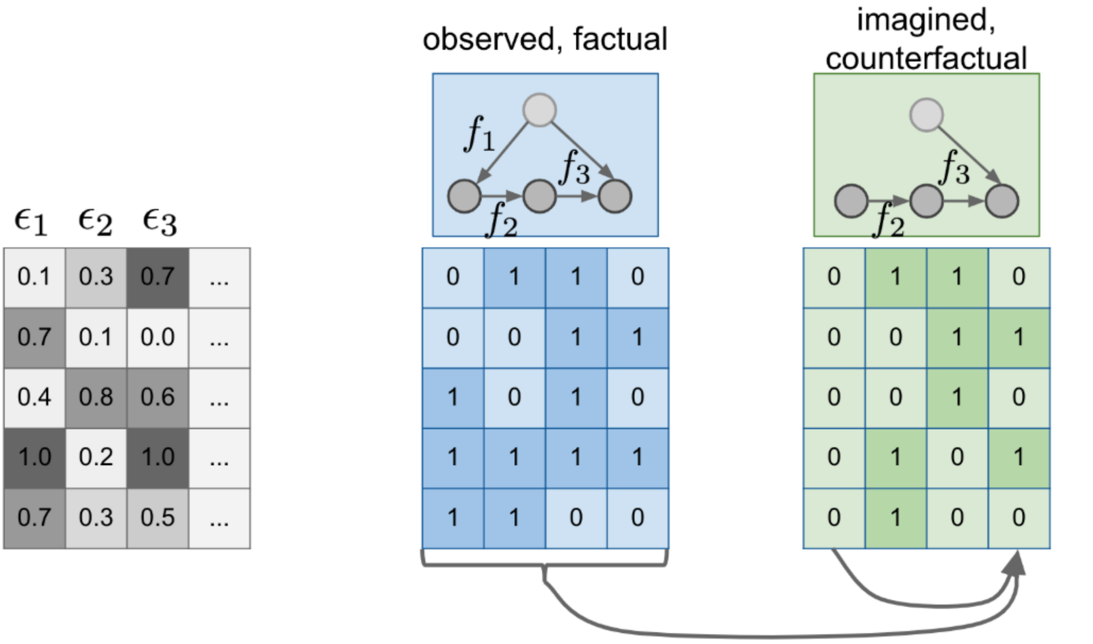
- I grab a guy randomly off the street and shave it's beard-- what is the probability of him getting a PHD? I expect do(beard=0) to be independent of PHD, so $P(PHD=1|do(Beard=0)) \doteq P(PHD=1)$
- Counterfactual focus on the probability of an outcome (PHD=1) of a specific PHD person had he did not get beard. Counterfactuals are "personalized" in the sense that you'd expect the answer to change if you substitute a different person in there.
- "My father has a beard and he does not have a PhD degree. I expect that preventing him to grow a mustache would not have made him any more likely to obtain a PhD. So his counterfactual probability would be a probability close to 0".

SEMs

- A--->B
- A and B are called "nodes".
- A is the parent of node B.



- Each variable/node in the graph has a corresponding function f.
- $x = f_1(u, e_1)$: u is a function of its "causal parent" u and an error term e_1.
- I can "reconstruct" the causal graph (the pic above) if I know $f_1 \ldots f_8$ (the system of equations of the SEM).
- In a SEM an intervention on a variable, say x, can be modelled by deleting the corresponding function, f_1, and replacing it with another function. For example $do(X = x_o)$ would correspond to a simple assignment to a constant $\tilde{f}_1(u) = x_o$.



ϵ_1	ϵ_2	ϵ_3	
0.1	0.3	0.7	...
0.7	0.1	0.0	...
0.4	0.8	0.6	...
1.0	0.2	1.0	...
0.7	0.3	0.5	...

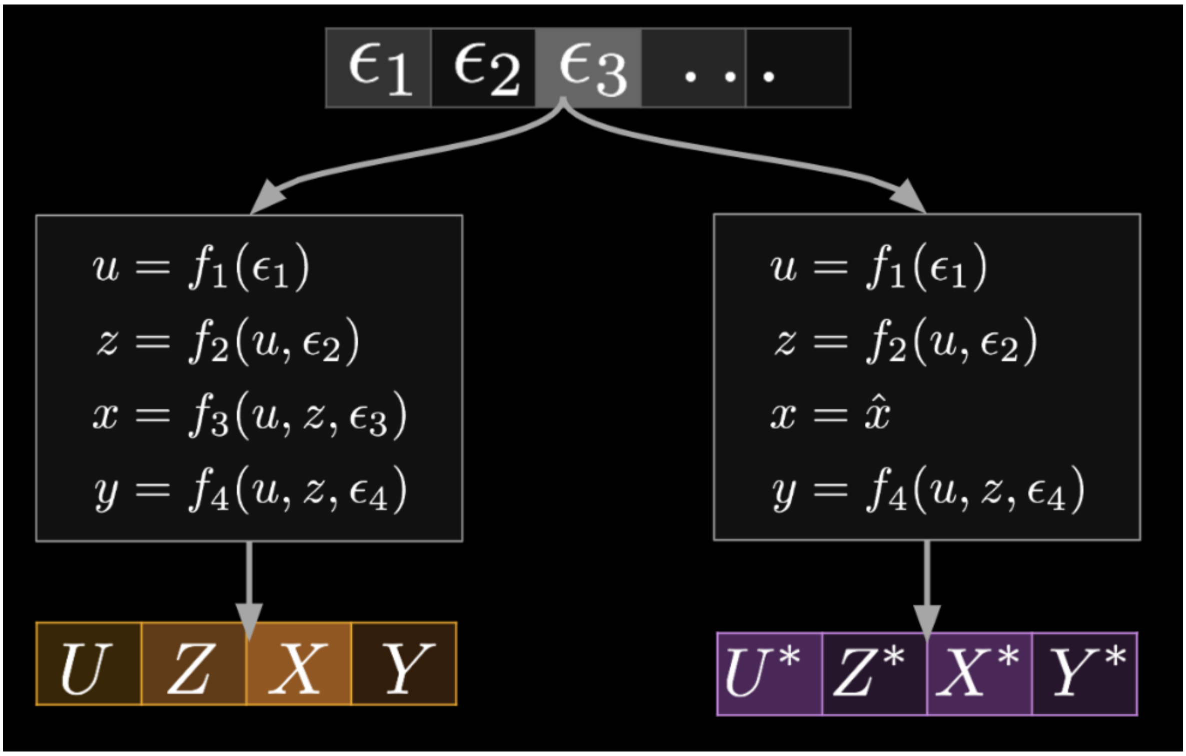
0	1	1	0
0	0	1	1
1	0	1	0
1	1	1	1
1	1	0	0

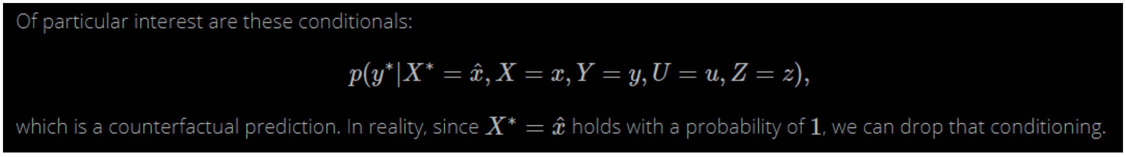
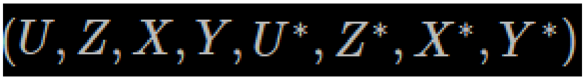
0	1	1	0
0	0	1	1
0	0	1	0
0	1	0	1
0	1	0	0

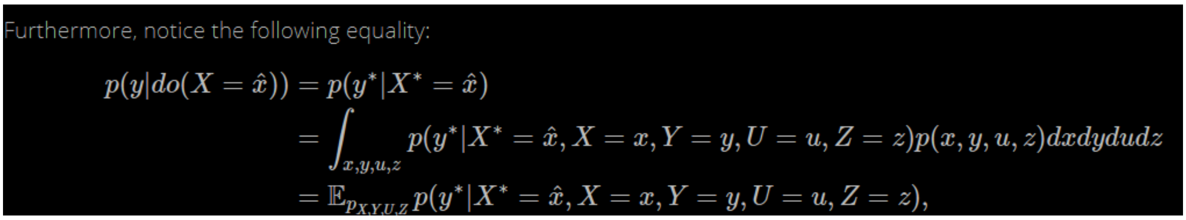
$p($🎓$^*|$🧔$^* = 0,$ 🧔 $= 1,$ 💍 $= 1,$ 💪 $= 1,$ 🎓 $= 1)$

- Counterfactuals are "making prediction" about features of the unobserved twin datapoint based on features of the observed datapoint.
- Because we used the same ϵs in both the blue and the green SEM. This induces a joint distribution between variables in the observable regime, and variables in the unobserved, counterfactual regime.

- Columns of the green table are no longer independent of columns of the blue table. You can start predicting values in the green table using values in the blue table, as illustrated by the arrows below them.
- Variables with an * are unobserved (and unobservable since) variables that live in the counterfactual world, while variables without * are observable.



- The RHS is the "parallel universe" where $x = \hat{x}$ (x is my intervention variable).
- 
  Of particular interest are these conditionals:
  $$p(y^*|X^* = \hat{x}, X = x, Y = y, U = u, Z = z),$$
  which is a counterfactual prediction. In reality, since $X^* = \hat{x}$ holds with a probability of 1, we can drop that conditioning.
- I find the above from the joint distribution over:
- 
  $(U, Z, X, Y, U^*, Z^*, X^*, Y^*)$
- Thus, you can find all sorts of conditionals and marginals from this joint probability.


Furthermore, notice the following equality:

$$\begin{aligned} p(y|do(X = \hat{x})) &= p(y^*|X^* = \hat{x}) \\ &= \int_{x,y,u,z} p(y^*|X^* = \hat{x}, X = x, Y = y, U = u, Z = z) p(x, y, u, z) dx dy du dz \\ &= \mathbb{E}_{p_{X,Y,U,Z}} p(y^*|X^* = \hat{x}, X = x, Y = y, U = u, Z = z), \end{aligned}$$

- The intervention conditional probability $P(Y|do(X = \hat{x}))$ is the average of counterfactuals over the observable population (which takes on all possible combinations of x,y,u,z).

- Author's summary: Counterfactual judgments remain hypothetical, subjective, untestable, unfalsifiable. You can't go back in time and rerun history with exactly the same circumstances except for a tiny change. But there are tools that exist to make predictions about counterfactuals.